



ORIGINAL ARTICLE

Assumptions, uncertainty, and catastrophic/existential risk: National risk assessments need improved methods and stakeholder engagement

Matt Boyd¹  | Nick Wilson² 

¹ Adapt Research Ltd, Reefton, New Zealand

² Department of Public Health, University of Otago, Wellington, Wellington, New Zealand

Correspondence

Matt Boyd, Adapt Research Ltd, 14 Broadway, Reefton 7830, New Zealand.

Email: matt@adaptresearchwriting.com

Abstract

Two key shortcomings of national risk assessments (NRAs) are: (1) lack of justification and transparency around important foundational assumptions of the process, (2) omission of almost all the largest scale risks. Using a demonstration set of risks, we illustrate how NRA process assumptions around time horizon, discount rate, scenario choice, and decision rule impact on risk characterization and therefore any subsequent ranking. We then identify a neglected set of large-scale risks that are seldom included in NRAs, namely global catastrophic risks and existential threats to humanity. Under a highly conservative approach that considers only simple probability and impact metrics, the use of significant discount rates, and harms only to those currently alive at the time, we find these risks have likely salience far greater than their omission from national risk registers might suggest. We highlight the substantial uncertainty inherent in NRAs and argue that this is reason for more engagement with stakeholders and experts. Widespread engagement with an informed public and experts would legitimize key assumptions, encourage critique of knowledge, and ease shortcomings of NRAs. We advocate for a deliberative public tool that can support informed two-way communication between stakeholders and governments. We outline the first component of such a tool for communication and exploration of risks and assumptions. The most important factors for an “all hazards” approach to NRA are ensuring license for key assumptions and that all the salient risks are included before proceeding to ranking of risks and considering resource allocation and value.

KEYWORDS

deliberative processes, existential risk, global catastrophic risk, national risk assessment, national risk register, public policy, stakeholder engagement

1 | INTRODUCTION

Many countries undertake the process of national risk assessment (NRA) to evaluate risks of national significance (OECD, 2017; Poljanšek et al., 2019). NRA often takes an all-hazards approach assessing natural hazards, infectious diseases, industrial accidents, terrorist attacks, labor strikes, cyberattacks, organized crime or the failure of institutions (OECD, 2017). The process is typically demanding, complex, multidisciplinary, and cross-sectoral.

NRA processes, such as those of the Netherlands (Veland et al., 2013) and the United Kingdom (Stock & Wentworth, 2019) involve development of risk scenarios of national significance, which are then evaluated in terms of their multicriteria impact and multicomponent likelihood. NRAs tend to exclude risks with low probability and have not usually been intended as an exhaustive register of all civil emergencies although lists of “risks under review” may be maintained, as in the United Kingdom (Stock & Wentworth, 2019), and Switzerland maintains a “Hazard Catalog” from which

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Risk Analysis* published by Wiley Periodicals LLC on behalf of Society for Risk Analysis.

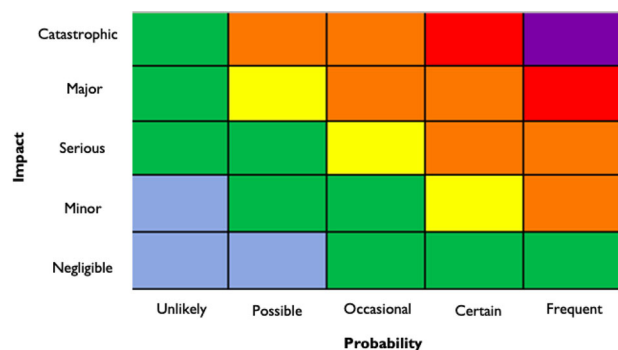


FIGURE 1 A probability-impact risk matrix. Darker colors (purple, red, orange) represent more salient risks than lighter colors (yellow, green, blue).

hazards are selected for further analysis in NRAs (FOCP, 2020). The focus is short-term (less than 5 years) so that resources are “not wasted” (OECD, 2017). A majority of OECD countries performing NRA communicate results in some form of national risk register (NRR) and/or consequence-probability (C,P) risk matrix (OECD, 2017). The risk matrix is a two-dimensional communication tool that “represents risk,” often while attempting to avoid spurious precision about probability or impact by placing risks in categories (see Figure 1).

However, this common practice of presenting a two-dimensional risk matrix often obscures uncertainties, stakeholder disagreements on values, bias, and systemic errors (Mamuji & Etkin, 2019). There may be lack of public engagement and awareness of the NRA/NRR (Hiscock & Jones, 2017), or the NRA itself is classified (Boyd & Wilson, 2021). Furthermore, recent analysis has argued that a consequence-uncertainty (C,U) definition of risk is more valid and useful when characterizing global and national risks than the (C,P) representation (Aven, 2017, 2020), and it is notable that uncertainty (and the strength of knowledge it is based upon) is not fully captured in many NRA/NRRs. NRA processes and the NRR outputs have been criticized on other grounds in a wide range of writings. We present a summary of these criticisms with supporting citations in Table 1.

NRA processes and the resulting NRAs/NRRs have multiple purposes. One aim of NRA should be to find common understanding across stakeholders of risks and priorities, effect debate, and achieve resilience to risk (Brody, 2020; Poljanšek et al., 2019). Although excluding localized risks, NRAs may help and inform local risk assessments and stimulate local authorities to build capacity and capability and prioritize resource allocation. The assessments can identify common consequences across multiple risks for which countries might prepare (Stock & Wentworth, 2019). A public summary of the NRA/NRR could raise awareness and nudge businesses and citizens to take self-protective measures (OECD, 2017) and support government actions.

Once the NRA has produced a set of risk evaluations (perhaps expressed as an NRR) prioritization of risks is possible. Prioritization is sometimes explicitly intended through the

NRA process in order to drive national resource prioritization decisions (Government Office for Science, 2012), or the “relative benefits of buying down risks” (OECD, 2009). In other cases it has been debated whether prioritization is a key function or not (Pruyt et al., 2013; Vlek, 2013b), and in yet other cases prioritization is explicitly not intended (Hagmann & Cavelti, 2012). Whether or not the NRAs/NRRs produced by the NRA process are presently used for prioritization, there is certainly a strong argument that they ought to be, particularly where new or overlooked risks and consequences are identified that may need to be managed. Methods for prioritization of national risks have been developed over the years, and various deliberative methods for risk ranking, involving stakeholder engagement, have been evaluated (Florig et al., 2001; Lundberg & Willis, 2016; H. H. Willis et al., 2004).

However, as we discuss below, such deliberative methods depend on the risks selected for inclusion, and summaries or characterizations of those risks. Risk characterizations in turn depend on foundational assumptions of the NRA process. Choices made when conducting NRAs will necessarily bias the results. Two key factors likely to have a large impact are:

1. The choice of key methodological assumptions, such as time horizon, discount rate, choice of scenario development (all of which logically precede deliberative methods for risk ranking), and decision rule choice (implicit in deliberative processes, but which could be made explicit).
2. Which risks are chosen for inclusion in the process (again preceding any deliberation comparing risks). At present most NRAs exclude very large-scale risks such as existential threats to humanity, and their “merely” catastrophic manifestations (global catastrophic risks [GCRs]).

The result of factors (1) and (2) is that NRAs and the resulting NRRs are only valid in a narrowly defined set of empirical and normative space and, as we demonstrate below, omit almost all the likely actual risk.

Such important national decisions as which risks to characterize and assess and what assumptions to base the assessment upon, are issues that might merit public consultation and transparent expert peer-review, especially where different value-, decision- or calculation-approaches preferred by stakeholders might lead to different process mechanics and ultimately different prioritization decisions. However, there is generally no mechanism for the public to explore the outputs of NRA under varying sets of process assumptions and included or excluded risks, and thereby provide expert-peer or democratic feedback.

1.1 | Aims

In this article we aim to demonstrate some shortcomings of existing NRA processes and outputs, namely: (1) how the choice of fundamental NRA process assumptions makes a material difference to the NRR output and any subsequent

TABLE 1 Some criticisms of national risk assessment processes and national risk register outputs.¹

Category of critique	Criticism	Sources
Foundational aspects	Incomplete/improper integration of societal values with the risk assessment and/or unclear decision rules	(Veland et al., 2013; Vlek, 2013a; Willis, Potoglou, de Bruin, & Hoorens, 2012)
	Limited debate and expert/stakeholder engagement/authorization/understanding	(Bossong & Hegemann, 2016; Government Office for Science, 2012; Hagmann & Cavelti, 2012; Hilton & Baylon, 2020; Hiscock & Jones, 2017; Lin, 2018; Stock & Wentworth, 2019; Vlek, 2013a)
	Lack of a standard (justified and effective) risk methodology	(Brody, 2020; Hagmann & Cavelti, 2012; Mamuji & Etkin, 2019; Stock & Wentworth, 2019)
	Methodological imprecision/confusion (e.g., around concepts of risk, probability, uncertainty, impact)	(Aven, 2020; Veland et al., 2013; Vlek, 2013a)
Risk and scenario selection	Not all salient risks are included or considered for inclusion (post-hoc inclusion of risks evident)	(Blagden, 2018; Deville & Guggenheim, 2018; Raine, 2021)
	Cognitive biases, groupthink, or institutional inertia obstruct valid assessment	(Blagden, 2018; Government Office for Science, 2012; Stock & Wentworth, 2019)
	Limitations of a single “reasonable worst-case scenario” approach, e.g., omission of decision-relevant information	(Bradley & Roussos, 2021; Hilton & Shah, 2021; Stock & Wentworth, 2019)
	Improper exclusion of uncertain, improbable, emerging, and devastating risks	(Etkin, Mamuji, & Clarke, 2018; Government Office for Science, 2012; Hilton & Baylon, 2020; Mamuji & Etkin, 2019)
	Over focus on risk within borders rather than potentially global consequences (nationalist bias)	(Hagmann & Cavelti, 2012)
	Time horizon excludes long-term risks (presentist bias)	(Stock & Wentworth, 2019)
	Imprecision and ambiguity of consequence/probability estimates	(Blagden, 2018)
Assessment	Spurious accuracy of consequence/probability estimates and/or lack of sensitivity analysis	(Government Office for Science, 2012; Hagmann & Cavelti, 2012)
	Insufficient accounting for interactions among risks and cascading effects (isolated hazard rather than integrated systems focus)	(Blagden, 2018; Gill & Malamud, 2016; Government Office for Science, 2012; Stock & Wentworth, 2019)
	Politicization of the risk assessment process and outcomes and/or vested interests	(Bossong & Hegemann, 2016; Brody, 2020; Deville & Guggenheim, 2018; Hagmann & Cavelti, 2012)
	Lack of focus on why things can go wrong and how policy contributes to the risk of things going wrong	(Hagmann & Cavelti, 2012)
	Risk of false positive and false negative results	(Vlek, 2013a, 2013b)
Outputs & validation	Lack of, or impossibility of, external validation	(Vlek, 2013b; Willis et al., 2012)
	False equivalence of rare devastating and common negligible risks	(Mamuji & Etkin, 2019)
	Risk assessments not effectively connected to risk management (not solution focused)	(Bossong & Hegemann, 2016; Lin, 2018; Raine, 2021)
	Circularity of scenarios being an outcome of policy choice, but scenarios assumed in choosing policy	(Bradley & Roussos, 2021)
	Lack of methodological process for situation awareness and warning of risks	(Raine, 2021)

¹The criticisms presented in this table are representative and are based on a focused, non-systematic review of the literature conducted by the authors in October 2021. Not every criticism applies to every NRA or NRR (some are better and more comprehensive than others). Additional criticism of risk matrices and risk prioritization exists independent of their use in NRAs.

deliberation on risk, (2) the weaknesses and ambiguity of risk matrices for communicating NRAs, (3) a major class of risks often neglected by NRA, and (4) the difficulties that uncertainty poses. We then suggest how those undertaking NRA could enter a productive dialog with stakeholders, supported by an interactive communication and engagement tool, to overcome some of these difficulties.

2 | IMPORTANT ASSUMPTIONS IN NRA

In this section we introduce a hypothetical set of risks to illustrate some key issues when undertaking NRA and when using NRAs and risk matrices to communicate national risk or inform prevention and mitigation. We assign these risks probabilities and consequences (impacts). We assume that methods for estimating probability and impact exist and are fit for purpose, given that NRA does in fact deduce such values:

The [United Kingdom (UK)] NRA identifies, assesses and prioritizes a range of representative risk scenarios that are considered challenging yet plausible manifestations of the wider risk they represent. These risks are then characterised on the basis of both likelihood and impact assessments (Government Office for Science, 2012).

Our following discussion is agnostic as to how probabilities and consequences are established. We assume that such methods account for the various components of the risk, namely hazard, vulnerability, and exposure, and that it is possible to estimate the impact a risk scenario(s) might have assuming business-as-usual.

2.1 | Probability, impact, time-horizon, and discounting

We begin with a set of point probabilities and point expected fatalities in a “challenging yet plausible” (henceforth “reasonable”) scenario for a demonstration set of six risks A–F. These “reasonable” scenarios are intended to represent serious but not worst-case manifestations of each hypothetical risk. Simplification to just one impact attribute and specific values, allows us to abstract from uncertainty and incommensurable variables for illustrative purposes. That said, it is possible to reduce multiple attributes to single variables as is commonly done when creating indices, undertaking health cost-utility analysis, and in producing composite risk impact scores through mathematical aggregation (Komenantova et al., 2014). The Norwegian NRA, for example, explicitly collapses multi-attribute impacts into a single measure (DSB, 2014, see Fig. 21 in that publication). In a different approach the Swiss NRA monetizes multi-attribute impacts into a single value (FOCP, 2020), as is common practice in health cost-utility analysis. In what follows we are

less concerned with how NRAs deal mathematically with the multi-attribute nature of risk. We will focus on prior issues that include timeframe of concern, discount rate, how scenarios are chosen, and decision rule applied to the results. It may be that collapsing consequences to one attribute is problematic, however, this is what many NRAs actually do when they produce a risk matrix.

The population of interest is taken to be the entire world, although impact could be crudely scaled to any particular country or region under some plausible weighting assumption. Table 2 catalogs these hypothetical risks, their probabilities and assumed impacts.

Inspecting Table 2 it is immediately apparent that risk E scenario has the largest impact (1 million lives at risk). This means that avoiding it would result in avoiding the worst possible outcome given this set of risks. Targeting prevention and mitigation to this risk would be a maximin decision strategy. Although risk E is improbable at first ($p = 0.0001$ in year 1), its probability rises to an aggregate probability across 50 years of 0.781, making it quite likely that it will occur. The consequence in expectation of risk E in year one is the third highest among these demonstration risks, containing only 0.1 times the risk of risk D (100 versus 1,000 expected deaths). However, if considered across 50 years, risk E contains 15.6 times as much expected impact as risk D (without discounting the future) or 8.7 times as much if future outcomes are discounted at 3% per annum. It is also striking to observe in Table 2 that although some risks have a low probability of occurring per annum, for example B (0.05), they are almost certain to occur repeatedly in a 50 year period (e.g., 2.5 times for B). Risk B is uncertain but realistically inevitable.

Table 3 presents the same risks A–F, but cataloged according to their maximum possible impact, a “worst-case” (tail risk) approach rather than a reasonable scenario approach. Worst case analysis may be preferred by some decisionmakers because “reasonable scenarios,” or alternatively, expected value across a probability distribution, could omit key decision-relevant information (e.g., worst cases, most probable cases). In Table 3 the probabilities are adjusted to reflect only the worst possible instance. This means Table 3 omits the harm from Table 2 (which reflects the more likely, though less impactful instances of these risks), though adding the expected impacts from Table 2 would make little difference to the expected impacts in Table 3.

Inspecting Table 3, it is apparent that across 50 years, risk D poses a greater threat in expectation than risk E (unlike in Table 2 when only the “reasonable” not “worst-case” scenario is considered). Note also that risk E is 78% as impactful as risk D across 50 years without discounting, but only 43% as impactful in expectation when 3% discounting is applied. Furthermore, when we apply 3% discounting, risk F changes from being the third most impactful risk in expectation (across 50 years without discounting), to being only the fourth most impactful, although it remains “inevitable” with 1.5 occurrences expected in 50 years.

The ordinal priority of the risks A–F based on estimated deaths under the various assumptions can be extracted from

TABLE 2 Plausibly “reasonable scenarios” of six major illustrative global risks A–F for use in demonstrating the importance of assumptions used in national risk assessments (NRAs)

Risk	Impact (i) (deaths) if the “reasonable scenario” occurs	Probability (P) of the “reasonable scenario” per annum	Impact in expectation ($P \times i$) in year one (deaths)	Number of occurrences expected in 50 years ($P \times 50$)	Aggregate impact across 50 years (deaths)	Aggregate impact across 50 years under 3% discounting**
A	75,000	0.001	75	0.05	3,750	1,930
B	1,000	0.05	50	2.5	2,500	1,290
C	10	0.2	2	10	100	51
D	10,000	0.1	1000	5	50,000	25,700
E	1,000,000	0.0001*	100	0.781 ⁺	781,000	223,000
F	10,000	0.05 [#]	500	3.5 ⁺	35,000	16,500

*Annual probability of risk E is nonstatic and increases fivefold each decade (to 0.00625 per annum in the fifth decade).

[#]Annual probability of risk F is nonstatic and increases by 0.01 each decade (to 0.09 per annum in the fifth decade).

⁺Reflects the aggregate probability accounting for the rising likelihood across 50 years.

**Discount rate of 3% used following Sanders et al. (2016) and rounded to three meaningful digits.

TABLE 3 Plausible “worst-case” scenarios of six major illustrative global risks A–F for use in demonstrating the importance of assumptions used in national risk assessments (NRAs)

Risk	Impact (i) (deaths) if “worst-case scenario” occurs	Probability (P) of worst-case scenario per annum	Impact in expectation ($P \times i$) in year one (deaths)	Number of occurrences of worst-case expected in 50 years ($P \times 50$)	Aggregate impact across 50 years (deaths)	Aggregate impact across 50 years under 3% discounting**
A	500,000,000	0.0001	50,000	0.005	2,500,000	1,290,000
B	100,000	0.01	1,000	0.5	50,000	25,700
C	1,000	0.1	100	5	5,000	2,570
D	100,000,000	0.02	2,000,000	1	100,000,000	51,500,000
E	1,000,000,000	0.00001*	10,000	0.0781 ⁺	78,100,000	22,300,000
F	2,000,000	0.01 [#]	20,000	1.5 ⁺	3,000,000	1,250,000

*Annual probability of risk E is nonstatic and increases fivefold each decade (to 0.00625 per annum in the fifth decade).

[#]Annual probability of risk F is nonstatic and increases by 0.01 each decade (to 0.09 per annum in the fifth decade).

⁺Reflects the aggregate probability accounting for the rising likelihood across 50 years.

**Discount rate of 3% used following Sanders et al. (2016) and rounded to three meaningful digits.

Table 2 and Table 3 across the four different evaluation methods displayed, that is, most impactful “reasonable scenario”, most impactful “worst-case” scenario, impact in expectation (annualized) considering only the next year, considering 50 years, with discounting, and without discounting. Priority of the top three risks under these decision assumptions is listed in Table 4.

The various evaluation assumptions lead to different priorities of risks across almost every evaluation. Two candidate risks vie for top priority (D and E), four risks vie for second priority (A,D,E,F) and four for third priority (A,D,E,F).

2.2 | Probability-impact matrices

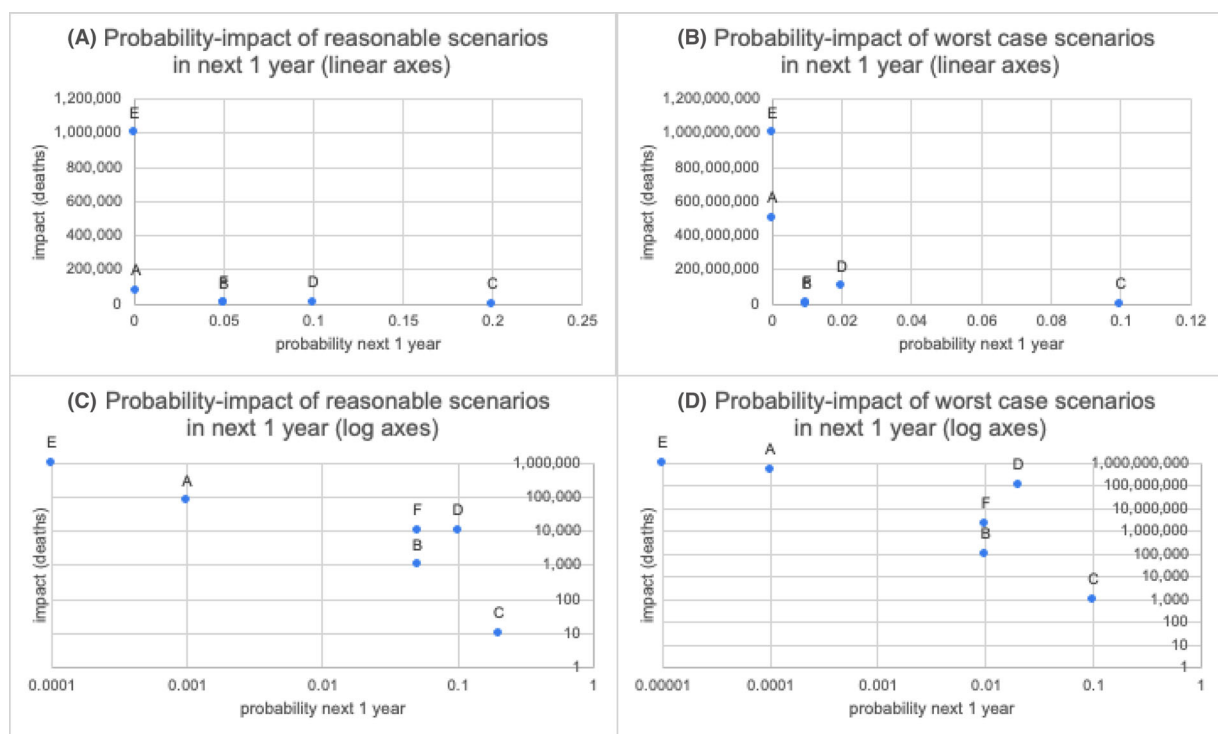
The use of probability-impact (risk) matrices has been thoughtfully criticized previously (Cox, 2008). However, it is useful to reiterate some of these weaknesses by demonstrating clearly how they appear to be of little help in untangling

the present examples. Risks A–F can be placed in a risk matrix as depicted in Figure 2.

Inspecting Figure 2, with Figure 1 in mind, it is apparent that probability-impact matrices are contingent, indeed fairly arbitrary, constructions. This is because decisions about the quantitative range of the axes and the choice of a linear or logarithmic scale, for example, will bias which risks fall into the risk categories (e.g., “catastrophic”, “major”, “serious”, etc.). Indeed, interpretations of terms like “unlikely” versus “possible” or “minor” versus “major” will differ by individual (Visschers et al., 2009). Figure 1 shows how color coding identifies these regions to a stakeholder, however, a simple adjustment to extend the x-axis all the way to probability 1.0 in Figure 2, panel B, would render risks B, D and C all very much in the “bottom left” (light blue) corner. The opposite effect ensues if one were to extend the x-axis of panel D leftwards substantially. What is a priority in absolute terms depends on a stakeholder’s quantitative frame and appetite for risk.

TABLE 4 Priority of illustrative risks when considering a range of evaluation methods (*ceteris paribus*—All other things being equal)

Evaluation method	Top priority	Second priority	Third priority
Most impactful “reasonable scenario” (ignoring probability) (see Table 2)	E	A	D & F
Most impactful “worst-case scenario” (ignoring probability) (see Table 3)	E	A	D
Impact in expectation next year under “reasonable scenario” assumption ($p \times i$) (see Table 2)	D	F	E
Impact in expectation next year under “worst-case scenario” assumption ($p \times i$) (see Table 3)	D	A	F
Impact in expectation aggregated across 50 years under “reasonable scenario” assumptions ($p \times i$) (see Table 2)	E	D	F
Impact in expectation aggregated across 50 years under “worst-case scenario” assumptions ($p \times i$) (see Table 3)	D	E	F
Discounted (3%) impact in expectation aggregated across 50 years under “reasonable scenario” assumptions ($p \times i$) (see Table 2)	E	D	F
Discounted (3%) impact in expectation aggregated across 50 years under “worst-case scenario” assumptions ($p \times i$) (see Table 3)	D	E	A

**FIGURE 2** Probability impact matrices for demonstration risks A–F with continuous, rather than categorical, linear, and logarithmic axes, across the next one year (“reasonable scenario” approach on the left, “worst-case” approach on the right).

There is a cluster of risks (F,D,B) in panel C of Figure 2, positioned toward the highest priority region of this risk matrix. As depicted, a stakeholder might imagine that these three risks have similar salience. However, from Table 2 we see that risk D has 20 times the impact in expectation in the next year compared to risk B. Logarithmic axes and especially categorization decisions mask much information, before even considering uncertainty about probability and consequences, and the strength of knowledge that underpins uncertainty (Aven, 2020), discussed below.

The four panels of Figure 2 all mask the salience of risk E. This risk is the highest priority risk under four of the

eight evaluation approaches in Table 4 (and second in the others), yet it consistently appears in the “green” region of these risk matrices (when overlaid with Figure 1). This positioning equates risk E with risk C (also “green”), which across 50 years will likely kill 100 people, compared to 781,000 for risk E. A number of decision rules would prioritize E over C, these include a standard maximization of expected utility rule, and also a maximin rule.

Figure 3 presents risk matrices containing risks A–F when considered across 50 years. Probability is given as the number of times a risk is expected to occur in this period (i.e., frequency). The impact is the consequence per occurrence.

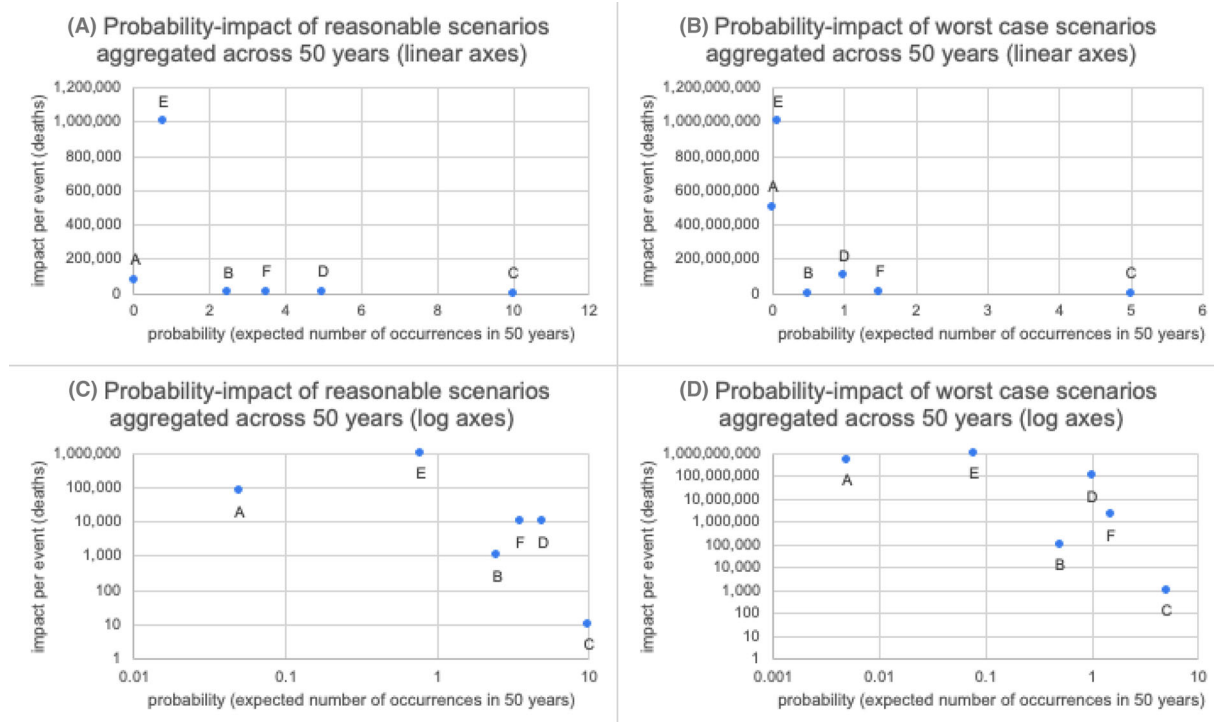


FIGURE 3 Probability impact matrices across 50 years for demonstration risks A–F, with continuous, rather than categorical, linear, and logarithmic axes (“reasonable scenario” approach on the left, “worst-case” approach on the right).

A notable feature in Figure 3 is that risks to the right of where $x = 1$, are almost certain to happen in the next 50 years. This has implications for the “foreseeability” of risks (beyond the common one-to-five-year time window of NRAs) and the imperative for prevention/mitigation. We know from Table 3 that risk E is 1,562 times more impactful in expectation than risk B across 50 years (78.1 million lives at risk vs 50,000). Figure 3 panel B communicates an inkling of that ratio, but it is rendered opaque in the logarithmic version (panel D) where risks E and B appear in the same moderate risk matrix “color” zone.

2.3 | Section summary

In Section 2 we focused on the probability and consequences by scenario choice, time horizon and discount rate. Even within just these relatively simple dimensions we illustrate that temporal scope and baseline assumptions make a material difference to how risks can be characterized. Additionally, the various risk matrices presented obscure important information contained in Table 2 and Table 3 rather than communicating it. The choice of a single risk matrix, as is common in NRAs, imposes a single set of opaque and value-laden assumptions. Sometimes, the assumptions and their rationale are detailed in NRA methodological documentation, however this information can be opaque or obscure from the perspective of many stakeholders, and alternative assumptions could be the focus of evidence-informed public debate. Furthermore, Table 4 shows how prioritization of

risks is not constant when scenario choice and timeframe of interest are varied. Any risk summaries developed for risks A–F to support deliberative ranking necessitate a priori agreement about these underlying assumptions of the NRA process.

Beyond the fundamental assumptions discussed above, three additional issues are critical to produce valid and useful NRAs: first, decisions around which risks are included, with particular emphasis on identifying the greatest threats, second, negotiating issues of uncertainty and third, effective communication of risks to stakeholders and the necessity of gathering public feedback and expert peer-review on methodological assumptions and deliberation in a structured way. Sections 3 and 4 address these additional issues.

3 | CATASTROPHIC AND EXISTENTIAL RISKS

Globally each year around 60,000 people are killed by “natural” disasters, including all geophysical, meteorological and acute climate events including earthquakes, volcanic activity, landslides, drought, wildfires, storms, and flooding (Ritchie & Roser, 2021). However, individual risks in Table 3 approximate (risk A) or vastly exceed (risk D) this total in expectation. The maximum impact of risk A is 500 million lives and for risk E it is 1 billion.

Most simply defined, global catastrophic risks can be considered to be those that could lead to the deaths of 10%,

TABLE 5 Illustrative list of some potential global catastrophic and existential risks

- Artificial intelligence (out-of-control and unaligned to human goals)
- Asteroid/comet impact
- Biological weapon use/accidental release
- Biotechnological advances (e.g., gain-of-function research, synthetic biology) resulting in pandemics
- Climate change
- Environmental degradation
- Extraterrestrial astrophysical processes, including coronal mass ejection
- Global food shortage
- Great power war
- Nanotechnology that goes awry
- Naturally occurring pandemic (e.g., influenza, coronavirus)
- Nuclear war and nuclear winter
- Nuclear terrorism
- Particle physics experiments that go awry
- Supervolcanic eruption
- Totalitarianism that becomes embedded
- Unforeseen anthropogenic risk (e.g., ozone depletion from chlorofluorocarbons was unexpected)
- Interactions or cascades of any of the above risks

Compiled from multiple sources: (Beard et al., 2020; Bostrom & Cirkovic, 2008; Ord, 2020), albeit with minor adaptations.

or more, of the current human population (Global Challenges Foundation, 2016). This is upwards of 800 million lives. Such risks may include the natural risks of extreme pandemic, supervolcanic eruption, asteroid or comet impact, or synchronous agricultural failure, among others. However, most of the catastrophic risk probably results from human activities, in particular technological advance. These anthropogenic catastrophes include nuclear war and nuclear winter, bioengineering and biological weapons, unaligned artificial intelligence (AI), climate change, and ecological collapse (Bostrom & Cirkovic, 2008; Ord, 2020). Existential risks are a limiting set of catastrophes that could lead to the premature extinction of humanity, or the permanent and drastic destruction of its potential (Ord, 2020). Existential risks are unprecedented and would not allow for meaningful recovery. We present a more complete list of these risks in Table 5.

Although we do not commit to A–F being modeled on any specific risks, they might plausibly represent:

- A. Supervolcanic eruption causing global cooling and agricultural failure (Jagermeyr et al., 2020; Rampino, 2008) (though Jagermeyr et al. model a regional nuclear conflict, it is relevant here as the impact on climate and agriculture is likely similar for certain kinds of volcanic eruption)
- B. Terrorism including nuclear terrorism (Ackerman & Potter, 2008)
- C. Industrial disasters (industry disasters surpassing 1,000 fatalities since the year 2000 include explosion, factory collapse, and ship sinking)
- D. Severe agricultural shortfall (e.g., due to synchronous drought, or disabled industry) (Denkenberger et al., 2017; Denkenberger & Pearce, 2016)
- E. Unaligned AI catastrophe (Beard et al., 2020; Bostrom, 2014; Russell, 2019)
- F. Biological threat such as an emerging infectious disease (Madhav et al., 2017; Marani et al., 2021)

Global catastrophic and existential risks are seldom included in NRAs and NRRs. For example, the NRR for the United Kingdom does not mention supervolcano, nuclear winter, or famine, and only mentions AI in passing (HM Government, 2020). Table 6 presents the NRAs for five countries and indicates whether they consider a representative set of eight GCRs or not.

Every national risk assessment differs, but none include all (or even many) of the GCRs. None contemplated potentially emerging catastrophic risks from AI or a biologically engineered pandemic. Most NRAs focused on influenza as the main natural pandemic risk, and most considered non-influenza natural pandemics to have impact far less than Covid-19 actually did. Only the Norwegian NRA hinted at catastrophic global climate impacts from large volcanic eruptions, but after mentioning, in a single sentence, cooling of the Earth by several degrees (possibly the single most catastrophic impact considered by any NRA due to the devastating food shortages implied and their cascading consequences), the issue was not revisited. Nuclear “attack” is commonly considered, but not nuclear war or winter. Only the Swiss NRA contemplates impact from near Earth objects, two of five NRAs do not mention solar flares/coronal mass ejections, and none identify global food shortage as a catastrophic risk.

Additionally, the consequences considered in these NRAs are certainly not “worst case”. For example, Norway designates “earthquake in a city” as the highest consequence event, indicating that the treatments of biological, nuclear, volcanic, and other risks, were very much not “worst case” scenarios, even though the examples we illustrated above show that “worst case” is where most of the expected consequence lies.

These omissions could be due to lack of awareness, uncertainty, or because probability-impact estimates, and other ex ante assumptions mean the risks fail to reach an inclusion threshold, or because policy options are thought to be limited. However, there is considerable understanding of, and concern over, these risks among a broad range of institutions that study them (Beard & Torres, 2020; CSER,

TABLE 6 Five illustrative national risk assessments (2011–2020).

Risk	UK (HM Government, 2020)	Switzerland (FOCP, 2020)	Netherlands (National Network of Safety and Security Analysts, 2019)	Norway (DSB, 2014)	New Zealand (DPMC 2011) ¹
Artificial intelligence	No	No	No	No	No
Pandemic from bioengineered agent	No	“biological attack” (not specifically engineered)	No	No	No
Natural pandemic	“Pandemics” (rated in most concerning category “E”) [however note: non-influenza infections were thought to threaten 100 lives in the 2017 version of the UK NRR]	“influenza pandemic” (but assessed as less damaging than “earthquake” and “electric power supply shortage”)	“Human infectious diseases” (specifically influenza, “bird flu”, ~10,000 fatalities) [however note 22,000 fatalities due to Covid-19 as at 5 July 2022]	“Pandemic in Norway” ~6,000 deaths (highest likelihood, but impact assessed as less than “nuclear accident”, also mostly focused on influenza)	“Human Pandemic” (rated as “catastrophic”)
Supervolcano	“Volcanic eruption” “C” (mostly focused on local ash, NRR does not examine global climate impact)	“Volcano eruption abroad” (~US\$1 billion impact in Switzerland)	No	“Long-term volcanic eruption in Iceland” (mentions “cooling of the earth by several degrees” but does not contemplate the catastrophe this implies - assessed as “moderate impact”)	“Very large eruption” (rated as “catastrophic”)
Nuclear war	“Major CBRN attack”—but the description implies a terrorist attack at some specific UK location “E”	“nuclear attack” (but not nuclear war or nuclear winter)	“military threats” and “proliferation of CBRN weapons” (but only mentions “low-yield nuclear weapons” and nuclear materials in the wrong hands)	“Strategic attack [on Norway]” (highest consequence, but similar to “earthquake in a city”)	“Global conflict” (rated as “major”, does not mention nuclear weapons)
Asteroid/comet	No	Yes “meteor strike”	No	No	No
Coronal mass ejection	Yes	Yes “solar storm”	No	Yes “100 year solar storm” Medium consequences	No
Major global food shortage	No	No	No	No	No

¹The New Zealand national risk assessment is classified as of 2022, however the results here were extracted from a 2011 discussion document on national security that presented a risk matrix of national risks.

2019; Global Challenges Foundation, 2016). Additionally, expected value estimates result in high priority within standard NRA methodology, even when considering only a 1-year time horizon, and especially when using longer horizons, maximin rules, and “worst-case” impact approaches (as demonstrated in Section 2). The *prima facie* case for including GCRs/existential risks in NRAs appears strong with the possibility of dominance across a range of attributes. Such

risks have little chance of being thoughtfully deliberated on if they are not characterized and included.

If we conservatively consider only the existential risks among the set of GCRs and ignore the more likely but nonexistential manifestations of the same risks, we see that several of these risks harbor consequences in expectation greater than all typically occurring natural hazards combined. In these cases, where impact is by definition “eight billion deaths” (or

TABLE 7 Risks of human existential catastrophe in the next 100 years (illustrative estimates for use in demonstrating the importance of assumptions used in national risk assessments [NRAs])

Risk	Probability in 100 years ¹	Probability per annum	Total deaths ²	Annualized deaths in expectation
Unaligned AI	0.1	0.001	8,000,000,000	8,000,000
Engineered pandemic	0.033	0.000333	8,000,000,000	2,670,000
Unforeseen anthropogenic disaster	0.033	0.000333	8,000,000,000	2,670,000
Other anthropogenic disaster	0.02	0.0002	8,000,000,000	1,600,000
Nuclear war	0.001	0.00001	8,000,000,000	80,000
Climate change	0.001	0.00001	8,000,000,000	80,000
Environmental damage	0.001	0.00001	8,000,000,000	80,000
All typically occurring natural disasters (for comparison)	1	1	60,000³	60,000
Supervolcanic eruption	0.0001	0.000001	8,000,000,000	8,000
Natural pandemic	0.0001	0.000001	8,000,000,000	8,000
Major asteroid/comet impact	0.000001	0.00000001	8,000,000,000	80
Stellar explosion	0.000000001	1E-11	8,000,000,000	0

¹Probabilities of existential catastrophe are those synthesized by Ord in his review of existential risks. These express an, “evidential sense of probability, which describes the appropriate degree of belief we should have on the basis of the available information”, and should be treated as “representing the right order of magnitude” (Ord, 2020). Additional published probability estimates for existential risks (employing multiple methodologies) have been reviewed by Beard et al. (2020) and could be substituted. Beard et al. (2020) note that 45% of the estimates they obtained relied at least in part on the subjective beliefs of the author, particularly with regard to AI risks, which “many Existential Risk scholars believe to be the most significant [risk]”.

²Considering the number of deaths due to an existential catastrophe to be 8 billion is a simplification. On one hand an existential catastrophe may permanently reduce the potential of humanity, without killing literally everyone, on the other hand 8 billion lives at risk does not consider the loss of future generations.

³Annual deaths from typically occurring natural disasters as estimated by Our World in Data (Ritchie & Roser, 2021). Note this excludes infectious diseases.

“drastic” curtailment of humanity’s potential) then expected value hinges only on probability (and its uncertainty).

Table 7 displays these risks as found in one recent systematic analysis, along with their probability of causing human existential catastrophe across a century (Ord, 2020). For additional estimates we direct the reader to a review of 67 published probability estimates across these risks (Beard et al., 2020). We have added “all typically occurring natural disasters” as a point of reference and extrapolated the annual probabilities and annualized deaths in expectation from these risks based on a population of eight billion people and assuming static probabilities. We are agnostic on whether these are the reasonable probabilities. Our point is to argue that such risks should be analyzed and considered in NRAs since not all these risks appear statistically unlikely. Additionally, should such risks be included in NRAs and NRRs, then this has implications for the prevention and management of risk with a more outward looking, cooperative global perspective perhaps being justified (e.g., a global risk assessment coordinated by the United Nations). Many of these risks might originate “elsewhere” but cascade to have catastrophic consequences for many, or every, country.

The annual probabilities for some of the risks in Table 7 will initially be lower than tabulated (the probability of existential catastrophe due to AI is almost certainly lower than 0.001 in 2023), but will rise as human activities generate additional risk, resulting in the aggregate 100 year probabilities, or until preventive and mitigation measures are in place. For this reason, we presented risks “E” and “F” above with rising probability. These probability trajectories under-

score the importance of timeframe selection, especially where mitigation might be a multidecade matter as with climate change.

Additionally, Table 7 focuses on existential risk only. Less impactful versions of the same risks are possible, and even very likely, such as pandemics or nuclear wars that do not kill every single person on earth. These must be considered when considering total risk (and different “levels” of impact could be disaggregated, with associated probabilities for each level—see Section 4). Table 7 also considers only presently existing people and excludes impacts on future generations in line with a “person-affecting” ethics (Parfit, 1984) contra to the rival “longtermist view” (Greaves & MacAskill, 2021). This is another conservative assumption of our analysis, and one that is contested. However, considering impact on the far future would only further increase the salience of existential risks.

The main takeaway from Table 7 should be that there are many risks harboring greater annualized impact in expectation (based on the probability of the outcome “human existential catastrophe”) than the sum of human lives lost from all typically occurring natural disasters. This is when only considering extinction cases, and only considering deaths of people alive at the time. We would not be surprised, once formal risk characterization is complete, to see some of these risks dominate, that is, score higher on consequence (across any attribute) and probability than many risks that governments currently substantially invest in mitigating.

Deliberation over such risks and whether they ought to be prioritized for mitigation, can only happen if they

are included in the NRA, characterized, communicated to stakeholders, and put forward to resource prioritization processes.

3.1 | Uncertainty about catastrophic and existential risks

The point estimates in the tables above, convey false precision. There is substantial uncertainty around most risks and this uncertainty has seldom been adequately reflected in assessments of large-scale risks (Aven, 2020; Aven & Cox, 2016; Veland et al., 2013). We can be uncertain about the probability of some risk and uncertain about the consequences. The degree to which we are certain depends on the strength of the knowledge underpinning our judgments (Aven, 2020). The strength of knowledge will depend on factors such as relevance of available data, degree of agreement among experts, thoroughness of assessment, and so on. It has been noted that two risks could be in the same position on a risk matrix, but the strength of knowledge that underpins each differs greatly (Aven, 2020).

Furthermore, the consequences of some hazards can manifest across a spectrum from trivial to devastating. It may be more helpful to consider there being different classes of a hazard (e.g., minor, moderate, severe, [existential]) which could be disaggregated and characterized independently. Contributors to the UK Lord's Report on Preparing for Extreme Risks have emphasized this point (House of Lords Select Committee, 2021). Disaggregation allows the most extreme (e.g., existential risk) or most probable (e.g., minor impact) manifestations of a hazard, to be represented independently. Such disaggregation has long been suggested in the literature on extreme events (Lambert, Matalas, Ling, Haimes, & Li, 1994). The Swiss NRA takes this approach, but then only compares the "major" (and not "significant" or "extreme") intensity profiles for each hazard (FOCP, 2020).

A practical point is that governments are often averse to investing under uncertainty. Consider risk "E" described above (perhaps equivalent to an unaligned AI catastrophe), the probability is low at first, and likely very uncertain, but it rises such that the cumulative risk becomes high over time. The decision when to act depends on the velocity of that rise. In such cases monitoring the change in probability is critical, and this monitoring (to increase strength of knowledge and reduce uncertainty) is the first step in risk prevention and mitigation. The need to invest in monitoring is not apparent if the NRA process looks only at years 1–5 where probability is initially low, with low consequence in expectation.

Uncertainty is inescapable in NRAs and it is important to characterize and scrutinize the relevant knowledge (Aven, 2020). Where there is high uncertainty but also high potential impact, then it may be wise to invest substantially in research and analysis (and wide consultation) to try and reduce the uncertainty. It might be possible to use public and expert engagement to help fill the gaps, increase knowledge, monitor changes in knowledge over time, identify warning signals

(especially for risks with rising probability), or refine probabilities, but crucially, this is only possible if the uncertain risks are included in the assessment and only if adequate resources are allocated to these tasks. One component of such scrutiny is public communication of the risk assessment and two-way engagement to source additional information and perspectives. We turn to this in Section 4.

4 | AN INTERACTIVE PUBLIC NATIONAL RISK COMMUNICATION AND ENGAGEMENT TOOL

Most NRA processes involve little public consultation, in some instances overt secrecy (Boyd & Wilson, 2021; OECD, 2017), and there is a documented lack of awareness of NRRs, even among local authorities to whom they are in part directed (Hiscock & Jones, 2017). This is despite the UN Global Assessment Report on Disaster Risk Reduction (2019) advocating, "increased access to risk information" and that, "low risk awareness is one of the main challenges" (United Nations, 2019).

Several arguments support public engagement tools for NRAs. For example, groupthink among those undertaking NRAs may lead to important omissions, perhaps due to political unwillingness to "contemplate unpalatable outcomes" (Blagden, 2018). Additionally, NRA processes can be politicized, and more scrutiny may be needed to counteract public and private authorities seeking to maintain their legitimacy (Blagden, 2018; Hagmann & Cavelti, 2012). Given the dependence of uncertainty upon knowledge, processes are needed to reduce surprise and scrutinize knowledge (Aven, 2020). Indeed, the UK House of Lords Report calls for "a more dynamic, data driven web-portal which allows users to visualise the risk summary, access the underlying data and easily navigate to related risks", furthermore, "risk identification exercises can benefit from broad 'crowd funding' input", and, "key players inside and outside Government should be asked for their input to challenge and sense-check the risk assessment" (House of Lords Select Committee, 2021).

As demonstrated above, NRAs and the resulting NRRs hinge upon a foundational set of assumptions and methods that determine the kinds of risks considered for inclusion and their characterizations. It is unclear how aware citizens are of these foundational assumptions, their implications, their alternatives, and how this might impact deliberation on risk prevention and mitigation resource allocation nationally, and internationally.

We note that scrutiny must logically first be applied to the underlying process assumptions, then to the resulting empirical claims, and finally deliberative prioritization (for mitigation or further research) can take place. We now propose the development of a freely available, open-access, risk communication and engagement tool to facilitate discussions on NRAs. Aspects of such a tool could be tailored to experts and other aspects to the general public.

4.1 | Rationale for wide expert engagement

Expert engagement, beyond the limited group of experts often consulted by government, is needed to scrutinize knowledge, identify omissions, provide evidence, overcome implicit biases, surface early signals or changing knowledge, and generally provide sensitivity analysis. The need for wider input is supported by examples of previous critical omissions and shortcomings of NRAs. These have included the omission of volcanic threats from the UK NRR prior to the eruption of the Icelandic volcano Eyjafjallajökull, or the assumed impact of “emerging infectious diseases” which, prior to Covid-19, was anticipated to be “up to 100 deaths” in the United Kingdom’s 2017 NRR (UK Cabinet Office, 2017). This is despite the coronavirus threat being known, as evidenced by the October 2019 “Event 201” coronavirus pandemic simulation exercise that modeled 65 million deaths (Center for Health Security, 2019). This example illustrates the need to decide whether the relevant scenarios analyzed in NRAs should be some “reasonably expected” case or the “worst case”, where most of the risk lies (as illustrated above).

In some domains wide expert engagement may increase the strength of knowledge and thereby reduce uncertainty. In other areas it might provide new information that emphasizes uncertainty and potentially draws attention to risks that had been previously dismissed. Both results are useful. Structured approaches, and especially enhanced solicitation techniques have been underdeveloped within the field of existential risk research (Beard et al., 2020). Therefore, as catastrophic and existential risks are considered for inclusion in NRAs, structured expert elicitation should occur in parallel.

4.2 | Rationale for wide public engagement

The 2020 Swiss NRA report details the type of contributors. Two-thirds (64%) were public sector employees, 26% represented “economy” (predominantly critical infrastructure operators), 10% were academics, and zero were public, business, non-government organization leaders, or others (FOCP, 2020). However, in a democracy decisionmakers should at least know what public preferences are even if they do not align with those of experts. While engagement with experts could well improve the quality and validity of NRAs, engagement with the public is an opportunity to inform and communicate risk and gather social license for foundational assumptions. Policy founded on widespread public input is more likely to persist across time, an important consideration for long-term risks.

Public-facing database- or model-driven interactive tools are available to help people understand the impact of varying assumptions or policy in other domains. For example, there are interactive Covid-19 policy tools that also represent uncertainty (Childs et al. n.d.). Although this Covid-19 example cited is hosted by a university, government could deploy similar information/engagement strategies. Our World

in Data’s many interactive tools (Our World in Data, 2022), or the World Health Organization’s Global Burden of Disease interactive platform visualizing mortality data (WHO, 2022), are other examples of publicly-facing interactive databases that can inform and prompt reflection. Such tools could support a public discussion on the assumptions and values that underpin NRAs.

Foundational assumptions and values upon which policy can be based have been elicited in jurisdictions such as Taiwan by using structured public engagement and online deliberation tools (the social media platform: Polis) (Tang, 2022). Public engagement could support more ethical, inclusive, and legitimate decision making. Such legitimacy might be particularly important when it comes to decisions around large, potentially multidecade or transgenerational investments in managing catastrophic risks. Importantly, the public needs to be informed and understand the range of assumptions about national risks in order to provide valid input.

4.3 | Engaging stakeholders and the public on national risks

The public has previously been engaged in risk ranking exercises based on validated deliberative methods for ranking risks in a range of settings and on both local and national risk issues (Florig et al., 2001; Lundberg & Willis, 2016; Willis et al., 2004). Participants reported satisfaction with the processes and their use for risk-management decisions (Willis et al., 2010). The deliberative process included eliciting which risk attributes must be covered in risk characterizations (Lundberg & Willis, 2016).

However, fundamental assumptions such as those about time horizon of the risk assessment, any temporal discount rate applied to consequences, and type of scenario choice (as discussed above) will frame how risk attributes (whichever are chosen) are characterized. A more fundamental engagement step is required prior to deliberative risk ranking to determine what is deliberated upon.

4.4 | An interactive national risk tool

To facilitate wider expert and public engagement with NRAs, we propose governments develop interactive risk communication tools that could allow users both internal and external to the public sector to explore the characterization of various risks under varying assumptions, including when global catastrophic and existential risks are considered alongside risks currently in NRAs. These tools should ideally be dynamic, web-based improvements on risk matrices with stakeholders able to alter assumptions and observe the new results.

The purpose of a new communication and engagement tool would be fourfold:

1. To provide information to the public and stakeholders on the probability and impact of national risks, and how these compare and rank ordinally and quantitatively and where uncertainty exists.
2. To enable interactive exploration of the ordinal rankings of national risks by varying fundamental assumptions including, but not necessarily limited to, timeframe, discount rate, scenario choices, decision rules, and decision-relevant risk attributes (see below).
3. To collect data on the preferred assumptions of stakeholders, independent experts, and the public, as well as collect submissions for additional risk classes, probability estimates and impact estimates along with up-to-date supporting evidence.
4. To provide access to the underlying risk data, encouraging interested users to consider all relevant information and allowing them to experiment with innovative methods for analysis, ranking, or visualization, thereby proposing improvements for consideration.

Such an interactive tool would require a supporting database. This could be as simple or as complex as resources and goals dictate. Ideally all major risks including global catastrophes, existential risks, and future risks (as listed in Table 5) should be included in the dataset so that they are not overlooked during deliberation. Users should be able to “test” the inclusion of new risks by completing a standard risk summary form.

The database should eventually include information on attributes of risks such as: an individual’s chance of death, the number of deaths possible (perhaps for minor, major and catastrophic versions of the risk), the number of (minor and major) illnesses/injuries, the time between exposure and harm, the degree of uncertainty (or strength of knowledge), and the ability of an individual to control their exposure. Included attributes should be based on the literature about which attributes influence risk preferences (Slovic et al., 1985), ideally validated in the relevant jurisdictional context.

4.5 | How such a tool might work

In practice a user would select their assumptions (guided by some screening questions, and perhaps facilitated by “slider” tools) around timeframe, discount rate, reasonable scenario(s) or worst-case impact, and decision rules. Options might even include others not discussed here such as impact on future generations. The user could identify which attributes to include in weighted aggregate impact variables as well as set attribute weights. The output would be a ranking of the consequence in expectation for each risk under those assumptions. Consequences could be represented as a single selected attribute, or as a variable that aggregates multiple weighted attributes.

The resulting ordinal ranking will avoid lossiness of decision relevant information or confusion when logarithmic

or categorical metrics are used to present risk matrices. Uncertainty bounds could be displayed where strength of knowledge is low, and users could visualize overlap and nonoverlap in expected consequences for risks under various assumptions. Users could include or exclude any of the risks in the database.

For interested experts, the tool would also enable exploration of the risk probabilities and expected consequences and the uncertainty around these data, by allowing the user to switch among point estimates, uncertainty bounds, and probability distributions.

The ordinal and quantitative rankings have several functions: first, they demonstrate the relative magnitude of risks, second, they invite reflection on the (public or government) user’s assumptions and the impact they have on the results, and third the rankings invite critique if users disagree with the quantitative ordering or if the user determines that key risks are absent. The tool would invite reasoned submission on such disagreements in the common language of the tool and could aggregate these alternate estimates into a crowd cognition result. A living database could operate with incomplete data and allow users to submit evidence to fill gaps over time.

4.6 | Outcomes

This dynamic interaction would facilitate nuanced communication of risk information from government to the public as well as preference and knowledge communication from the public, stakeholders, and experts to government on designated assumptions, quantitative assertions, inaccuracies, and omissions.

In the future, quantitative estimates (of probability or consequences for example) could be uploaded by users similarly to prediction markets and users could choose to see aggregated “crowd” data, or only data sanctioned following rigorous government analysis. Pairwise comparisons could be elicited from users about which risks are more or less salient to inform risk rankings (and median weights implicitly applied to various risk attributes might be deduced across the population). Ideas for effective interventions to mitigate risks could be crowdsourced. The platform could contain links to tutorials that could better prepare people for engagement with government on NRAs. The creative possibilities for a tool like this and its potential to inform and support public conversations are very broad and it could be used creatively to explore some of the potential weaknesses of NRAs listed in Table 1.

5 | DISCUSSION

Important assumptions need to be agreed before a valid NRA can proceed. These include methodological and normative choices that determine which risks are included, and how these risks are characterized (across time). The most logical

next step is to characterize the risks including expressions of uncertainty. Risks can then be ranked according to some criteria for prioritization around mitigation or need for further research. Finally, rankings can be examined, along with information on the cost-effectiveness of interventions to lower or mitigate risks and other ethical, pragmatic, and political values in order to allocate resources. Our analysis has focused largely on the first step and we have argued that key assumptions make a material difference to the NRAs and that most current NRAs omit consideration of a suite of very important large scale risks (global catastrophic and existential risks).

We have highlighted the need to express uncertainty in risk communication products and that transparent engagement with the public and a broad array of experts supported by a two-way risk communication tool could: (1) help to legitimize key assumptions, (2) ensure omission of important risks is avoided (especially risks that might dominate existing risks even if this is uncertain), and (3) provide robust critique of risk characterizations and the knowledge underpinning them.

The short time-horizon of current NRAs appears particularly problematic. The probability of harmful outcomes from several anthropogenic risks is likely rising. Current NRA approaches may not be adequate for anticipating these rising probabilities, deploying resource to reduce uncertainty, and prevent and/or mitigate these threats in time. Given failures to act in domains such as climate change, the public might prefer such risks enter the assessment much earlier, input could be gathered about warning signs, and specified “fire alarms” might be needed. Any trade-off between near-term risk mitigation and long-term risk mitigation should be communicated publicly and quantitatively and be open to public debate.

Importantly, if one favors a long-term view of risk, a low discount rate, and a preference for decisions based on worst-case scenarios, then emerging risks pertaining to AI, biological risks, conflict, or ecological catastrophe may be more salient. We do not know to what degree the public/stakeholders take this view, because the options (time-frame of concern, value of the future, maximin versus other decision rules, and so on) have not been systematically put to them in the context of NRAs. We have suggested how this process might be initiated using a tool to overcome the secrecy that surrounds NRRs in some countries, uses existing NRA methods for probability and impact assessment, and that communicates uncertainty.

If one takes standard government cost-effectiveness analysis (CEA) as the starting point, especially the domain of healthcare where cost-per-quality-adjusted-life-year is typically the currency and discount rates of around 3% are typically used, then existential risk just looks like a limiting case for CEA. The population at risk is simply all those alive at the time and the clear salience of existential risks emerges in simple consequence calculations (such as those demonstrated above) coupled with standard cost-utility metrics. The cost-effectiveness of early action to understand, prevent and mitigate might be particularly high for risks such

as those that threaten global food security (Denkenberger & Pearce, 2016; Rivers et al., 2022) or health security (Millet & Snyder-Beattie, 2017).

The real question then becomes, why do government NRAs and CEAs not account for the probabilities and impacts of GCRs and existential risk? Possibilities include unfamiliarity (i.e., a knowledge gap, to be solved by wider consultation), apparent intractability (i.e., a lack of policy response options, to be solved by wider consultation), conscious neglect (due to low probability or for political purposes, but surely to be authorized by wider consultation), or seeing some issues as global rather than national (typically requiring a global coordination mechanism). Most paths point toward the need for informed public and stakeholder dialog.

Decisions over whether to examine risks in terms of scenarios that are challenging but reasonable to plan for, whether to assess the limiting “worst-case” for a particular risk, whether to aggregate total risk for a particular hazard by sampling from a probability-impact distribution, or whether to disaggregate the spectrum of risk into scenarios of different severity, are fundamental to risk assessment. For many risks, tail threats contain almost all the risk. On an expected utility approach to risk this cannot reasonably be ignored. We should not confound the plausibility of any particular scenario with the likely aggregate impact across all scenarios for a risk.

Realistic and concerning tail risks include disruptions to global agriculture (climate change, synchronous drought, nuclear winter, and major volcanic eruptions), severe pandemics, and major technological risks. It is striking that NRRs do not specifically include food shortages as a major risk (Hilton & Shah, 2021), nor many of the technological risks in Table 5. The probability of severe impacts may be low or uncertain today but could rise rapidly. We have argued that high uncertainty is not a reason to exclude risks from assessment. Additional research to scrutinize knowledge and reduce such uncertainty for the most salient risks can then be prioritized.

Concerns we identify above around groupthink of government experts, lack of transparency and awareness, potential politicization, and lack of legitimacy favor the development of interactive consultation tools that can support two-way discourse, informing the public and stakeholders about NRAs while at the same time gathering information on preferences and expert knowledge. We have not developed the tool we propose above (it will take some dedicated resources to do this), but we hope that the act of articulating this solution highlights the shortcomings of present centralized and closed approaches to national risk.

Finally, in this article we have focused on a few key normative and methodological factors that must be agreed before national risk characterization can be accomplished. These factors deal with establishing the “consequence in expectation (with uncertainty)” across a disparate suite of risks. Given the limitations of two-dimensional risk matrices, we favor a single-dimensional ranking of these consequences in

expectation especially for very large risks where linear representation would highlight the extreme salience of some risks. This approach could support a true “all hazards” analysis across government, acknowledging that “hazards” might better be analyzed in the form of “systems risks” (Avin et al., 2018).

6 | CONCLUSIONS

We identified two key shortcomings of NRA processes: (1) lack of justification and transparency around the foundational assumptions of the process, (2) the omission of almost all the largest scale risks. We then used a demonstration set of risks to illustrate the impact that choices about time horizon, discount rate, and the method for estimating the impact of a risk (e.g., reasonable scenario versus limiting worst-case) have on risk characterization. We identified a set of large-scale risks that are seldom included in NRAs and highlighted a range of uncertainties.

We illustrated the extreme salience of global catastrophic and existential risks through a highly conservative approach that considers only simple probability and impact metrics, the use of common discount rates, and harms only to those currently alive at the time. These assumptions are agnostic to some of the more esoteric normative frames advocated by some as foundations for the extreme importance of global catastrophic and existential risks. The salience of existential risks is possibly much higher than their omission from NRAs indicates, apparently vastly exceeding the salience of all typically occurring natural disasters combined. We have also suggested that standard CEA might reveal highly cost-effective measures to take against these extreme risks. Much greater focus may be warranted on risks from unaligned AI, pandemics/biosafety, global food shortages, or indeed any risks that expert assessment considers “existential” or even “catastrophic” (see Table 5).

We advocate the need for a deliberative public tool to support informed two-way communication between the public, stakeholders, and government, and in this article have outlined the first component of such a tool. Eventually, a wide-ranging interactive online tool should be able to communicate and support exploration of what are the most salient risks, where society/societies could feasibly act, and why should we act. Additionally, understanding the potentially catastrophic local impact of major global risks originating “elsewhere” might help integrate the findings of disparate NRAs into a globally cooperative approach.

The most important factor, however, for an “all hazards” approach is to make sure that all the salient risks are included before proceeding to these additional deliberations across risks, resource allocation and value. We have demonstrated the strong case that all realistic catastrophic and existential threats should be included in the process. Finally, NRA processes should ultimately connect globally and support international cooperation on risk assessment and mitigation.

ACKNOWLEDGMENTS

Open access publishing facilitated by University of Otago, as part of the Wiley - University of Otago agreement via the Council of Australian University Librarians.

FUNDING

Self-funded.

CONFLICT OF INTERESTS

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data used in this study are freely available online and sources are cited. Data comprising Tables 2 and 3 were devised by the authors for illustrative purposes.

ORCID

Matt Boyd  <https://orcid.org/0000-0002-1387-5047>

Nick Wilson  <https://orcid.org/0000-0002-5118-0676>

REFERENCES

- Ackerman, G., & Potter, W. (2008). Catastrophic nuclear terrorism: A preventable peril. In N. Bostrom & M. Cirkovic (Eds.), *Global catastrophic risks* (pp. 402–449). Oxford University Press.
- Aven, T. (2017). Improving risk characterisations in practical situations by highlighting knowledge aspects, with applications to risk matrices. *Reliability Engineering & System Safety*, 167, 42–48. <https://doi.org/10.1016/j.res.2017.05.006>
- Aven, T. (2020). How to determine the largest global and national risks: Review and discussion. *Reliability Engineering & System Safety*, 199, 106905. <https://doi.org/10.1016/j.res.2020.106905>
- Aven, T., & Cox, L. A. Jr. (2016). National and global risk studies: How can the field of risk analysis contribute? *Risk Analysis*, 36(2), 186–190. <https://doi.org/10.1111/risa.12584>
- Avin, S., Wintle, C., Weitzdorfer, J., O'hEigeartaigh, S., Sutherland, W., & Rees, M. (2018). Classifying global catastrophic risks. *Futures*, 102, 20–26. <https://doi.org/10.1016/j.futures.2018.02.001>
- Beard, S., Rowe, T., & Fox, J. (2020). An analysis and evaluation of methods currently used to quantify the likelihood of existential hazards. *Futures*, 115, 102469. <https://doi.org/10.1016/j.futures.2019.102469>
- Beard, S., & Torres, P. (2020). *Identifying and assessing the drivers of global catastrophic risk: A review and proposal for the global challenges foundation*. <https://globalchallenges.org/assessing-the-drivers-of-global-catastrophic-risk-final/>
- Blagden, D. (2018). The flawed promise of National Security Risk Assessment: nine lessons from the British approach. *Intelligence and National Security*, 33(5), 716–736. <https://doi.org/10.1080/02684527.2018.1449366>
- Bossong, R., & Hegemann, H. (2016). EU internal security governance and national risk assessments: towards a common technocratic model? *European Politics and Society*, 17, 226–241. <https://doi.org/10.1080/23745118.2016.1120990>
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Bostrom, N., & Cirkovic, M. Eds.. (2008). *Global catastrophic risks*. Oxford University Press.
- Boyd, M., & Wilson, N. (2021). Anticipatory governance for preventing and mitigating catastrophic and existential risks. *Policy Quarterly*, 17(4), 20–31. <https://doi.org/10.26686/pq.v17i4.7313>
- Bradley, R., & Roussos, J. (2021). Following the science: Pandemic policy making and reasonable worst-case scenarios. *LSE Public Policy Review*, 1(4), 6. <https://doi.org/10.31389/lseppr.23>

- Brody, M. (2020). Enhancing the organization of the United States department of homeland security to account for national risk. *Homeland and Security Affairs*, 16, Article 3. <https://www.hsaj.org/articles/15847>
- Center for Health Security. (2019). *Event 201*. <https://www.centerforhealthsecurity.org/event201/>
- Childs, M., Kain, M., Kirk, D., Harris, M., Ritchie, J., Couper, L., Delwel, I., Nova, N., & Mordecai, E. (n.d.). *Potential long-term intervention strategies for COVID-19*. Retrieved from <https://covid-measures.stanford.edu/>
- Cox, T. (2008). What's wrong with risk matrices? *Risk Analysis*, 28(2), 497–512. <https://doi.org/10.1111/j.1539-6924.2008.01030.x>
- CSER. (2019). *Managing global catastrophic risks Part 1: Understand*. <https://www.cser.ac.uk/resources/policy-series-managing-global-catastrophic-risks-part-1-understand/>
- Denkenberger, D., Cole, D., Abdelkhalik, M., Griswold, M., Hundley, A., & Pearce, J. (2017). Feeding everyone if the sun is obscured and industry is disabled. *International Journal of Disaster Risk Reduction*, 21, 284–290.
- Denkenberger, D., & Pearce, J. (2016). Cost-effectiveness of interventions for alternate food to address agricultural catastrophes globally. *International Journal of Disaster Risk Science*, 7, 205–215. <https://doi.org/10.1007/s13753-016-0097-2>
- Deville, J., & Guggenheim, M. (2018). From preparedness to risk: from the singular risk of nuclear war to the plurality of all hazards. *The British Journal of Sociology*, 69(3), 799–824.
- DPMC. (2011). New Zealand's National Security System. Wellington: New Zealand Department of Prime Minister and Cabinet. Retrieved from <https://dpmc.govt.nz/sites/default/files/2017-03/national-security-system.pdf>
- DSB. (2014). *National risk analysis 2014*. https://www.dsb.no/globalassets/dokumenter/rapporter/nrb_2014_english.pdf
- Etkin, D., Mamuji, A., & Clarke, L. (2018). Disaster risk analysis part 1: The importance of including rare events. *Journal of Homeland Security and Emergency Management*, 15(2), 20170007. <https://doi.org/10.1515/jhsem-2017-0007>
- Florig, H. K., Morgan, M. G., Morgan, K. M., Jenni, K. E., Fischhoff, B., Fischbeck, P. S., & DeKay, M. L. (2001). A deliberative method for ranking risks (I): Overview and test bed development. *Risk Analysis*, 21(5), 913–921. <https://doi.org/10.1111/0272-4332.215161>
- FOCP. (2020). *Disasters and emergencies in Switzerland 2020: National risk analysis report*. <https://www.babs.admin.ch/en/aufgabenbabs/gefaehdrisiken/natgefaehrdanalyse.html>
- Gill, J., & Malamud, B. D. (2016). Hazard interactions and interaction networks (cascades) within multi-hazard methodologies. *Earth System Dynamics*, 7, 659–679. <https://doi.org/10.5194/esd-7-659-2016>
- Global Challenges Foundation. (2016). *Global catastrophic risks report 2016*. <http://globalprioritiesproject.org/2016/04/global-catastrophic-risks-2016/>
- Government Office for Science. (2012). *Blackett review of high impact low probability risks*. <https://www.gov.uk/government/publications/high-impact-low-probability-risks-blackett-review>
- Greaves, H., & MacAskill, W. (2021). *The case for strong longtermism*. <https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/>
- Hagmann, J., & Cavelty, M. (2012). National risk registers: Security scientism and the propagation of permanent insecurity. *Security Dialogue*, 43(1), 79–96. <https://doi.org/10.1177/0967010611430436>
- Hilton, S., & Baylon, C. (2020). *Risk management in the UK: What can we learn from COVID-19 and are we prepared for the next disaster?* <https://www.cser.ac.uk/resources/risk-management-uk/>
- Hilton, S., & Shah, S. (2021). Why are food shortages not listed as a risk in the national risk register? <https://www.cser.ac.uk/resources/food-shortages-NRR/>
- Hiscock, K., & Jones, A. (2017). Assessing the extent to which the UK's national risk register supports local risk management. *Sustainability*, 9(11), 1991. <https://www.mdpi.com/2071-1050/9/11/1991>
- HM Government. (2020). *National risk register: 2020 Edition*. <https://www.gov.uk/government/publications/national-risk-register-2020>
- House of Lords Select Committee. (2021). *Select committee on risk assessment and risk planning report of session 2021–22: Preparing for extreme risks: Building a resilient society*. <https://publications.parliament.uk/pa/ld5802/ldselect/ldrisk/110/110.pdf>
- Jagermeyr, J., Robock, A., Elliott, J., Muller, C., Xia, L., Khabarov, N., Folberth, C., Schmid, E., Liu, W., Zabel, F., Rabin, S. S., Puma, M. J., Heslin, A., Franke, J., Foster, I., Asseng, S., Bardeen, C. G., Toon, O. B., & Schmid, E. (2020). A regional nuclear conflict would compromise global food security. *Proceedings of the National Academy of Sciences*, 117(13), 7071–7081. <https://doi.org/10.1073/pnas.1919049117>
- Komendantova, N., Mrzyglocki, R., Mignan, A., Khazai, B., Wenzel, F., Patt, A., & Fleming, K. (2014). Multi-hazard and multi-risk decision-support tools as a part of participatory risk governance: feedback from civil protection stakeholders. *International Journal of Disaster Risk Reduction*, 8, 50–67. <https://doi.org/10.1016/j.ijdr.2013.12.006>
- Lambert, J. H., Matalas, N. C., Ling, C. W., Haimes, Y. Y., & Li, D. (1994). Selection of probability distributions in characterizing risk of extreme events. *Risk Analysis*, 14(5), 731–742. <https://doi.org/10.1111/j.1539-6924.1994.tb00283.x>
- Lin, L. (2018). Integrating a national risk assessment into a disaster risk management system: Process and practice. *International Journal of Disaster Risk Reduction*, 27, 625–631. <https://doi.org/10.1016/j.ijdr.2017.08.004>
- Lundberg, R., & Willis, H. (2016). Deliberative risk ranking to inform homeland security strategic planning. *Journal of Homeland Security and Emergency Management*, 13(1), 3–33. <https://doi.org/10.1515/jhsem-2015-0065>
- Madhav, N., Oppenheim, B., Gallivan, M., Mulembakani, P., Rubin, E., & Wolfe, N. (2017). Pandemics: Risks, impacts, and mitigation. In D. T. Jamison, H. Gelband, S. Horton, P. Jha, R. Laxminarayan, C. N. Mock, & R. Nugent (Eds.), *Disease control priorities: Improving health and reducing poverty*. The International Bank for Reconstruction and Development /The World Bank. pp. 315–346.
- Mamuji, A., & Etkin, D. (2019). Disaster risk analysis part 2: The systemic underestimation of risk. *Journal of Homeland Security and Emergency Management*, 16(1), 20170006. <https://doi.org/10.1515/jhsem-2017-0006>
- Marani, M., Katul, G. G., Pan, W. K., & Parolari, A. J. (2021). Intensity and frequency of extreme novel epidemics. *Proceedings of the National Academy of Sciences*, 118(35), e2105482118. <https://doi.org/10.1073/pnas.2105482118>
- Millet, P., & Snyder-Beattie, A. (2017). Existential risk and cost-effective biosecurity. *Health Security*, 15(4), 373–383. <https://doi.org/10.1089/hs.2017.0028>
- OECD. (2009). *OECD Studies in risk management: Innovation in country risk management*. <https://www.oecd.org/futures/Innovation%20in%20Country%20Risk%20Management%202009.pdf>
- OECD. (2017). *National risk assessments: A cross country perspective*. https://www.oecd-ilibrary.org/governance/national-risk-assessments_9789264287532-en
- Ord, T. (2020). *The precipice: Existential risk and the future of humanity*. Bloomsbury.
- Our World in Data. (2022). *Research and data to make progress against the world's largest problems*. <https://ourworldindata.org/>
- Parfit, D. (1984). *Reasons and persons*. Clarendon Press.
- Poljanšek, K., Casajus Valles, A., Marin Ferrer, M., De Jager, A., Dottori, F., Galbusera, L., Garcia Puerta, B., Giannopoulos, G., Girgin, S., Hernandez Ceballos, M., Iurlaro, G., Karlos, V., Krausmann, E., Larcher, M., Lequarre, A., Theodoridou, M., Montero Prieto, M., Naumann, G., Necci, A., ... Wood, M. (2019). *Recommendations for national risk assessment for disaster risk management in EU*. <https://publications.jrc.ec.europa.eu/repository/handle/JRC114650>
- Pruyt, E., Wijnmalen, D. J., & Böklerink, M. (2013). What can we learn from the evaluation of the Dutch national risk assessment? *Risk Analysis*, 33(8), 1385–1388. <https://doi.org/10.1111/risa.12096>
- Raine, S. (2021). *Half of the national risk register is missing*. <https://rusi.org/explore-our-research/publications/rusi-newsbrief/half-national-risk-register-missing/>

- Rampino, M. (2008). Super-volcanism and other geophysical processes of catastrophic import. In N. Bostrom & M. Cirkovic (Eds.), *Global catastrophic risks* (pp. 205–221). Oxford University Press.
- Ritchie, H., & Roser, M. (2021). *Natural disasters*. <https://ourworldindata.org/natural-disasters#natural-disasters-kill-on-average-60-000-people-per-year-and-are-responsible-for-0-1-of-global-deaths>
- Rivers, M., Hinge, M., Garcia Martinez, J., Tieman, R., Jaeck, V., Butt, T., & Denkeberger, D. (2022). *Deployment of resilient foods can greatly reduce famine in an abrupt sunlight reduction scenario*. Research Square. <https://doi.org/10.21203/rs.3.rs-1446444/v1>
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Allen Lane.
- Sanders, G. D., Neumann, P. J., Basu, A., Brock, D. W., Feeny, D., Krahn, M., Kuntz, K. M., Meltzer, D. O., Owens, D. K., Prosser, L. A., Salomon, J. A., Sculpher, M. J., Trikalinos, T. A., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (2016). Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second panel on cost-effectiveness in health and medicine. *Journal of the American Medical Association*, 316(10), 1093–1103. <https://doi.org/10.1001/jama.2016.12195>
- Slovic, P., Fischhoff, B., & Lichtenstein, S. (1985). Characterizing perceived risk. In R. W. Kates, C. Hohenemser & J. X. Kaspersen (Eds.), *Perilous progress: Managing the hazards of technology*. Westview. pp. 91–125.
- Stock, M., & Wentworth, J. (2019). *Evaluating UK natural hazards: the national risk assessment*. <https://researchbriefings.files.parliament.uk/documents/POST-PB-0031/POST-PB-0031.pdf>
- Tang, A. (2022, 2 February) Audrey Tang on what we can learn from Taiwan's experiments with how to do democracy/Interviewer: R. Wiblin. 80,000 Hours Podcast.
- UK Cabinet Office. (2017). *National risk register of civil emergencies: 2017 edition*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/644968/UK_National_Risk_Register_2017.pdf
- United Nations. (2019). *Global assessment report on disaster risk reduction*. <https://gar.undrr.org/report-2019>
- Veland, H., Amundrud, Ø., & Aven, T. (2013). Foundational issues in relation to national risk assessment methodologies. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 227(3), 348–358. <https://doi.org/10.1177/1748006x12472870>
- Visschers, V., Meertens, R., Passchier, W., & De Vries, N. (2009). Probability information in risk communication: A review of the research literature. *Risk Analysis*, 29(2), 267–287. <https://doi.org/10.1111/j.1539-6924.2008.01137.x>
- Vlek, C. (2013a). How solid is the Dutch (and the British) national risk assessment? Overview and decision-theoretic evaluation. *Risk Analysis*, 33(6), 948–971. <https://doi.org/10.1111/risa.12052>
- Vlek, C. (2013b). What can national risk assessors learn from decision theorists and psychologists? *Risk Analysis*, 33(8), 1389–1393. <https://doi.org/10.1111/risa.12097>
- WHO. (2022). *WHO Mortality database: Interactive platform visualizing mortality data*. <https://platform.who.int/mortality>
- Willis, H., Potoglou, D., de Bruin, W., & Hoorens, S. (2012). *The validity of the preference profiles used for evaluating impacts in the Dutch national risk assessment*. https://www.rand.org/pubs/technical_reports/TR1278.html
- Willis, H. H., DeKay, M. L., Morgan, M. G., Florig, H. K., & Fischbeck, P. S. (2004). Ecological risk ranking: development and evaluation of a method for improving public participation in environmental decision making. *Risk Analysis*, 24(2), 363–378. <https://doi.org/10.1111/j.0272-4332.2004.00438.x>
- Willis, H. H., Gibson, J. M., Shih, R. A., Geschwind, S., Olmstead, S., Hu, J., Curtright, A. E., Cecchine, G., & Moore, M. (2010). Prioritizing environmental health risks in the UAE. *Risk Analysis*, 30(12), 1842–1856. <https://doi.org/10.1111/j.1539-6924.2010.01463.x>

How to cite this article: Boyd, M., & Wilson, N. (2023). Assumptions, uncertainty, and catastrophic/existential risk: National risk assessments need improved methods and stakeholder engagement. *Risk Analysis*, 43, 2486–2502. <https://doi.org/10.1111/risa.14123>